Testimony of Daniel A. Reed
Director, Renaissance Computing Institute
University of North Carolina at Chapel Hill

May 13, 2004

Good morning, Chairman Boehlert and Members of the Committee. Thank you very much for granting me this opportunity to comment on appropriate paths for scientific computing. I am Daniel Reed, Director of the Renaissance Computing Institute (RENCI), a collaborative activity of the University of North Carolina at Chapel Hill, Duke University and North Carolina State University. I am the former Director of the National Center for Supercomputing Applications (NCSA) at the University of Illinois, one of three NSF-funded high-end computing centers. I am also a researcher in high-performance computing.

In response to your questions regarding the High-Performance Computing Revitalization Act of 2004, I would like to make three points today regarding high-performance computing.

1. International Competitiveness

High-performance computing has emerged as the third element of the research portfolio, complementing theory and experiment. Computing breathes life into the underlying mathematics of theoretical models, allowing us to understand nuanced predictions and to shape experiments more efficiently. Computing also allows us to capture and analyze the torrent of experimental data being produced by a new generation of scientific instruments and sensors, themselves made possible by advances in computing and microelectronics.

Legend says that Archimedes remarked, on the discovery of the lever, "Give me a place to stand, and I can move the world." Today, computing pervades all aspects of science and engineering. "Science" and "computational science" have become largely synonymous, and high-performance computing is the intellectual lever that helps assure U.S. scientific leadership in an increasingly competitive world.

High-performance computing plays a special and important role as an intellectual lever by allowing researchers and practitioners to bring to life theoretical models of phenomena when economics or other constraints preclude experimentation. Computational cosmology, which tests competing theories of the universe's origins by computationally evolving cosmological models, is one such example. Given our inability to conduct cosmological experiments (we cannot create variants of the current universe and observe its evolution), computational simulation is the only feasible way to conduct experiments.

High-performance computing also enables researchers to evaluate larger or more complex models and to manage larger volumes of data than would be possible on conventional computer systems. Although this may seem prosaic, the practical difference between obtaining results in hours, rather than weeks or years, is substantial – it qualitatively changes the range of studies one can conduct. For example, climate change studies, which simulate thousands of Earth years, are only feasible if the time to simulate a year of climate in a few hours. Moreover, conducting parameter studies (e.g., to assess sensitivity to different conditions such as the rate of fluorocarbon or CO_2 emissions) is only possible if the time required for each simulation is small.

Finally, high-performance computing allows us to couple models to understand the interplay of processes across interdisciplinary boundaries. Understanding the environmental and biological bases of respiratory disease or biological attack requires coupling of fluid dynamics models to model airflow and inhalants, whether smoke, allergens or pathogens, materials models to surface properties and interactions, biophysics models of cilia and their movements for ejecting foreign materials, and deep biological models of the genetic susceptibility to disease. The complexity of these interdisciplinary models is such that they can only be evaluated using high-performance computers.

The breadth of these examples highlights a unique aspect of high-performance computing that distinguishes it from other scientific instruments – its universality as an intellectual amplifier. Powerful new telescopes advance astronomy, but not materials science. Powerful new particle accelerators advance high energy physics, but not genetics. In contrast, high-performance computing advances all of science and engineering, because all disciplines benefit from high-resolution model predictions, theoretical validations and experimental data analysis. As new scientific discoveries increasingly lie at the interstices of traditional disciplines, high-performance computing is the research integration enabler.

Although this universality is the intellectual cornerstone of high-performance computing, it is also its political weakness. Because all research domains benefit from high-performance computing, but none is solely defined by it, high-performance computing lacks the cohesive, well-organized scientific community of advocates found in other disciplines. In turn, this has led to over-dependence on market forces to shape the design and development of high-performance computing systems, to our current detriment.

Fueled by weapons research and national security concerns, until the 1980s, the U.S. government's high-performance computing needs could substantively influence the commercial market and assure U.S. supremacy in high-performance computing. Scientific and government high-performance computing needs are now a much smaller fraction of the overall computing market, with concomitantly less economic influence.

With the explosive growth of the computing industry and the internationalization of information technology, we are in danger of losing our international competitive advantage in high-performance computing, with serious consequences for scientific research and industrial competitiveness. This economic milieu has had profound effects on all aspects of high-performance computing – research and development, marketing, procurement and operation.

This brings me to my second point: the current status of our efforts.

2. Current Status and Coordinated Solutions

Not only has high-performance computing enriched and empowered scientific discovery, as part of a larger information technology ecosystem, it has also been responsible for substantial economic growth in the United States. **Because of this success, information technology and high-performance computing are increasingly international activities, with associated competition for intellectual talent and access to world-class computing resources**.

In an era of constrained Federal budgets and fierce international competition, we cannot afford wasted or duplicative efforts. The great strength of the U.S. research system is its diversity – many research ideas can be explored, with funding opportunities at multiple agencies. In computing, this diversity also creates leaks in the pipeline from basic research to deployment and commercial infrastructure, and many promising ideas are lost. The pipeline from basic research, through advanced prototyping and evaluation, to either research infrastructure or commercial development, requires tactical and strategic coordination across agencies.

Hence, we must encourage cross-agency collaboration and coordination, while leveraging the unique missions and attributes of each agency. Only via such inter-agency coordination can we maintain international leadership in high-performance computing. This belief is supported by broad community consensus. **During the past three years, at least six community reports[1] have highlighted the limitations of current approaches and have recommended an integrated, interagency initiative in high-performance computing.**

I applaud the Committee for capturing the central elements of these recommendations in the High-Performance Computing Revitalization Act, namely the need to (a) train a new generation of high-performance computing users and researchers, (b) conduct basic research and advanced prototyping for high-performance computing, and (c) develop and deploy high-performance systems that match scientific needs. **In addition to these goals, I recommend that the HPC Act also include mechanisms to aid the transfer of promising technologies to commercial practice.** The substantial engineering costs to develop high-performance computing systems and their limited market means that government incentives or support may prove necessary to sustain development of high-performance systems that can meet national scientific and security needs.

I believe an interagency initiative in high-performance computing should be based on the following principles

1. *An integrated strategic plan that articulates the responsibilities, scope and financial scale of each agency's responsibilities.*
2. *Regular deployment and support of the world's highest performance computing facilities for open scientific use, as part of a broad ecosystem of supporting infrastructure, including high-speed networks, large-scale data archives, scientific instruments and integrated software.*
3. *Coordination and support for national priorities in science, engineering, national security and economic competitiveness.*
4. *Vendor engagement to ensure technology transfer and economic leverage*
5. *Verifiable metrics of interagency collaboration, community engagement and technical progress that are tied to agency funding*

The National Science Foundation (NSF) and the Department of Energy (DOE) are the primary supporters of physical science and engineering research, whereas the National Institutes of Health (NIH) fund the majority of life science and biomedical research. Each of these and other Federal civilian science agencies has a unique, though critical role in the computing technology pipeline.

There has been much debate about the relative roles of NSF and DOE in providing access to high-performance computing for scientific research. This debate misses the critical point – the coordinated actions of both agencies are needed to ensure U.S. competitiveness, and both should be charged with deploying and operating systems with the highest possible capability.

- [1] NSF Workshop on Computation as a Tool for Discovery in Physics, September 2001
  - www.nsf.gov/pubs/2002/nsf02176
- Report on High-Performance Computing for the National Security Community, July 2002
  - www.hpcc.gov/hecrtf-outreach/bibliography/200302_hec.pdf
- Blueprint for Future Science Middleware and Grid Research and Infrastructure, August 2002
  - www.nsf-middleware.org/MAGIC/default.htm
- Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, January 2003
  - http://www.cise.nsf.gov/sci/reports/toc.cfm
- DOE Science Networking Challenge, June 2003
  - gate.hep.anl.gov/may/ScienceNetworkingWorkshop/
- DOE Science Case for Large Scale Simulation, June 2003
  - www.pnl.gov/scales/
- Community Workshop on the Roadmap for the Revitalization of High End Computing, June 2003
  - www.hpcc.gov/hecrtf-outreach

Reflecting its role as a basic research agency, NSF should support advanced systems research, including new architectures, software and tools and advanced algorithms. This research is the well spring of tomorrow's computing systems and infrastructure and the educational opportunity for a new generation of high-performance computing researchers. **Concurrently, NSF should continue to develop and support leading edge computing and data management systems, both for open community access and to support its Major Research Equipment (MRE) projects.**

Investments in "computing as science" (i.e., basic research in next generation computing technologies) and "computing for science" (i.e., deployment of computing infrastructure as a scientific enabler) are complementary, with qualitatively different time scales and needs. Given the rapid rates of change in computing technologies, high-performance computing infrastructure must be sustained at adequate levels for long periods and renewed regularly if it is to remain relevant to research facilities that have 10-20 year operational lifetimes.

Many high-performance computing research ideas can only be validated by constructing large-scale prototypes. In the 1970s and 1980s, the U.S. funded several research and development efforts in high-performance computing, and we continue to harvest insights from these experiments. Today, there are few, if any such projects, with concomitant loss of experience and insight. Hence, **DOE should lead advanced prototyping and deployment of next-generation high-performance computing systems, coupled to its scientific facilities and laboratory mission. This advanced prototyping and development should harvest basic research ideas from the DOE and NSF portfolios for national deployment.**

Finally, as quantitative biology and biomedicine expand to include tools and techniques from the physical and mathematical sciences, the National Institutes of Health (NIH) must also assume a leadership role in computational science and high-performance computing. **The biological research triumphs of the past decade were due in no small measure to a combination of biological insight and judicious application of new computing technology. Equally importantly, the biomedical discoveries of this decade, with concomitant cost savings and improved treatments, will depend critically on the deep integration of biology, medicine, software, algorithms and hardware. Hence, NIH should also lead by supporting both computing research and the creation of a national infrastructure for biomedical data sharing, computational modeling and distributed collaboration that is interoperable with that being deployed by NSF and DOE.**

While we debate appropriate actions, our international competitors are moving ahead. As part of the Sixth Framework, the European Union plans to deploy a pan-European Grid as a baseline infrastructure in support of scientific research. In the U.S., we are developing a set of loosely connected Grids without a common framework or strategic funding plan. Similarly, Japanese investment in the Earth System Simulator, the world's fastest computing system, is well known.

This leads me to my third and final point: research needs and opportunities.

3. Actions

The explosive growth of commodity clusters has reshaped the high-performance computing market. Although this democratization of high-performance computing has had many salutatory effects, including broad access to commodity clusters across laboratories and universities, it is not without its negatives. Not all applications map efficiently to the cluster programming model of loosely coupled, message-based communication, and it is difficult for vendors to make a profit developing systems tailored for scientific research. Hence, some researchers and their applications have suffered due to lack of access to more tightly coupled supercomputing systems. Second, **an excessive focus on peak performance at low cost has limited research into new architectures, programming models, system software and algorithms.**

The result has been the emergence of a high-performance "monoculture" composed predominantly of commodity clusters and small symmetric multiprocessors (SMPs).

In the 1990s, the U.S. high-performance computing and communications (HPCC) program supported the development of several new computer systems. In retrospect, we did not recognize the critical importance of long-term, balanced investment in hardware, software, algorithms and applications. Achieving high-performance for complex scientific applications requires a judicious match of computer architecture, system software, tailored algorithms and software development tools. We have substantially under-invested in the research needed to develop a new generation of architectures, programming systems and algorithms. The result is a paucity of new approaches to managing the increasing disparity between processor speeds and memory access times (the so-called von Neumann bottleneck).

Hence, **we must target exploration of new systems that better support the irregular memory access patterns common in scientific and national defense applications. In turn, promising ideas must be realized as advanced prototypes that can be validated with scientific codes.** In addition, we must recognize that new programming models and tools are needed that simplify application development and maintenance. The current complexity of application development unnecessarily constrains use of high-performance computing, particularly for commercial use. Finally, increases in achieved performance over the past twenty years have been due to both hardware advances and algorithmic improvements; we must continue to invest in basic algorithms research. This critical cycle of prototyping, assessment, development and deployment must be a long-term, sustaining investment, not a one time, crash program.

Opportunities abound for application of high-performance computing in both science and industrial sectors. Integrated vehicle designs with lifetime warranties, based on coupled electrical, mechanical and power train models, are within reach. Higher resolution cosmological models would allow testing of competing theories of the evolution of the universe, with sufficient resolution to simulate galaxy formation. Personalized medicines, tailored to minimize toxicity and maximize efficacy based on individual genetics, are possible based on drug chemistry models. All require a new generation of high-performance computing systems that can deliver high sustained performance for a suite of coupled models.

There is no "silver bullet" that will eliminate current problems and ensure continued U.S. preeminence in high-performance computing. Rather, the challenge is creating and sustaining an integrated, interagency research, development and deployment program that is reflective of national needs and opportunities. **Today, high-performance computing is reaping the rewards of yesterday's research investment. We must seed tomorrow's crop of research ideas today, else tomorrow we will subsist on wild berries**.

In conclusion, Mr. Chairman, let me thank you for this committee's longstanding support for scientific discovery and innovation. Thank you very much for your time and attention. I would be pleased to answer any questions you might have.

Biographical Sketch

Professor Daniel A. Reed is Director of the Renaissance Computing Institute (RENCI), an interdisciplinary center spanning the University of North Carolina at Chapel Hill, Duke University and North Carolina State University. He was previously Director of the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign, where he also led National Computational Science Alliance, a consortium of roughly fifty academic institutions and national laboratories that is developing next-generation software infrastructure of scientific computing. He was also one of the principal investigators and chief architect for the NSF TeraGrid. Professor Reed is also the former head of the Department of Computer Science at the University of Illinois, one of the oldest and most highly ranked computer science departments in the country. He holds the Chancellor's Eminent Professorship at the University of North Carolina at Chapel Hill where he conducts interdisciplinary research in high-performance computing.